International Journal of Scientific & Engineering Research, Volume 7, Issue 2, February-2016
ISSN 2229-5518

Product Opinion based on Review taken from Internet using Sentiment Analysis

Swati M.Chandurkar, Dr.M.A.Pund

Abstract— Sentiment analysis is the type of text classification according to sentiments present in text. Sentiment analysis includes identification of how sentiments are expressed in texts and whether the expressions indicate positive or negative opinions toward the subject. An important part of our information-gathering behavior has always been to find out what other people think. Online shopping sites encourage users for posting review about product they purchase. Such reviews are advantageous for new user for taking the decision about product at the time of purchasing and also for manufacturers of that product to take decision about its production. Therefore sentiment analysis of product review is becoming popular in text mining and computational linguistic research. This paper focuses on the categorization of a plain input text to inform a Text-to-speech system about the most appropriate sentiment and also on taking the review of real time product using sentiment analysis.

Index Terms— Text-to-speech (TTS) synthesis, feature engineering, sentiment analysis, text classification.

1 Introduction

Sentiment analysis of product reviews has recently become very popular in text mining. Sentiment Analysis means identification and classification of the opinion expressed in the text using information retrieval and computational linguistics. Opinion is necessary in our information gathering behavior before taking a decision. The internet is a valuable place for reviews for a product or service. There may be lot of reviews in the internet for a product or services therefore it is difficult to understand customer opinions. Online review sites, and personal blogs assists for gathering of sentiments of products or object using information technologies.
Sentiment analysis includes identification of

- Expressions of sentiments,
- Polarity and strength of the expressions, and
- Their relationship to the subject

To address the problem of sentiment classification, different common features that are of use to denote the affect in text should be described. In order to obtain them, a pipeline design pattern framework of Emolib is developed to label affect in text [15], [16].

IJSER staff will edit and complete the final formatting of your paper.

2 Related Work

The identification of affect in text is a dificult problem. Because of the prediction of affect in text, the attention of the text to speech (TTS) synthesis research community get attracted. The conventional sentiment analysis solutions may need to be adapted to the TTS environment as they are usually set to work with compilations of long texts that are not analyzed at sentence-level [18].

Lina Zhou et al, investigated movie review mining using machine learning and semantic orientation [20] and in the proposed machine learning approach, Supervised classification and text classification techniques are used for classifying the movie review and the study says that the supervised machine learning is more efficient but requires a considerable amount of time to train the model.

Bo Pang et al., used machine learning techniques for investigating the effectiveness of classification of documents by overall sentiment [21]. Experiments concluded that the machine learning techniques are better than human produced baseline for sentiment analysis on movie review data.

Zhu et al., proposed aspect based opinion polling from free form textual customers reviews [22]. The aspect related terms used for aspect identification was learnt using a multi-aspect bootstrapping method.

Jeonghee Yi et al., proposed a Sentiment Analyzer for extracting opinions about a subject from online data documents [23]. Sentiment analyzer uses natural language processing techniques.

Alekh Agarwal et al., proposed a machine learning method incorporating linguistic knowledge gathered through synonymy graphs, for effective opinion classification [24]. This approach shows the degree of influence among relationships of documents have on their sentiment analysis.

Michael et al., presented, a prototype system for mining topics and sentiment orientation from free text customer feedback [25].

Qui et al., analyzed the problems related to opinion mining such as opinion lexicon expansion and opinion target extraction [26]. Opinion targets are entities and their attributes on which opinions have been expressed.

Lei Zhang et al., [27] identified domain dependent opinion words. Noun and noun phrases that indicate the product feature which implies opinions are found using a feature based opinion mining model.

- *Swati M.Chandurkar is currently pursuing masters degree program.*
- *Dr.M.APund is professor at PRMIT&R Badnera,Amravati.*

It is dificult to translate human affect into explicit representations. One option is to presuppose the existence of some suitable taxonomy of affective states [6]. Other works, instead, delved into the relevant characteristics of the available text of analysis without enlarging the data to process [5].

3 Sentiment Analysis For Product Opinion

A. Framework

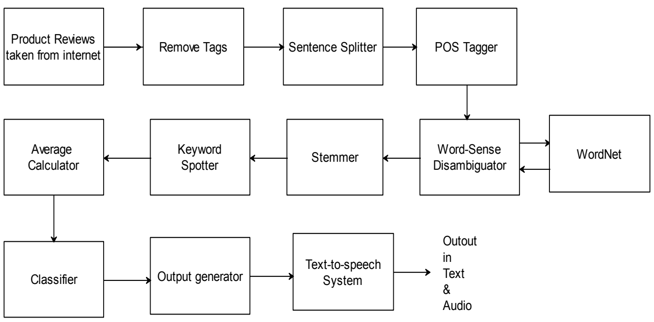

Fig 1: Pipeline framework for sentiment analysis of product review

As The framework for sentiment analysis of product review consists of following different modules:

1) Reviews of product

The input to the system will be reviews of product taken from related websites using Internet. As input is a real time input, it is dynamic in nature.

2) Remove Tags

In the real time reviews taken from internet may contain non-textual data and markup tags for html pages. Such information are not required for sentiment analysis and thus should be removed.

3) POS Tagger

Part-of-speech (POS) information is important in sentiment analysis and opinion mining. Part-of-speech tagging is refered as a crude form of word sense disambiguation. The POS Tagger is used to explain the significance of the words in the document. The adverbs, adjectives and verbs have significant effect on the sentence. Thus this module helps in identifying the most affective word in the given document. From this block the adverbs, adjectives and verbs are then used for the further processing.

4) Word-Sense Disambiguator

Human language is ambiguous, so that many words can be interpreted in multiple ways depending on the context in which they occur. For instance, consider the following sentences:
(a) She can hear bass sounds.
(b) She like grilled bass.

Here the word bass in the two sentences has different meanings:
Tones of low-frequency and a kind of fish, respectively. Many times humans do not think about the ambiguities of language, machines required to process unstructured textual information and transform them into data structures which must be analyzed for determining the underlying meaning. The identification of meaning for words in context is called word sense disambiguation (WSD).

5) Stemmer

It is used to remove the inflection of words for indexing. Semantically related words get mapped to the same stem, base or root form. Stemming is the process for reducing inflected words to their word stem, base or root form. Stemming programs is also called as stemming algorithms or stemmers. A stemming algorithm reduces the words ‘fishing’, ‘fished’, and ‘fisher’ to the root word, ‘fish’. On the other hand, ‘argue’, ‘argued’, ‘argues’, ‘arguing’, and ‘argus’ reduce to the stem ‘argu’.

6) Keyword Spotter

This module is used for developing monograms and bigrams. It takes up three features where initially the word features are identified by confirming whether or not the word exists in the document. It is like that of the next feature to identify the unigrams in the word and hence check for their availability. It provides the emotional dimensions to the emotional words.

7) Average calculator

It is used to compute the averaged emotional dimensions for the text of analysis. In the current work, this is the arithmetic mean of the dimensions at the sentence level [9], [11].

8) Classifier

It is used for prediction of the most appropriate sentiment label according to the features extracted from the text. The classifier then outputs the class that the input text belongs to either positive or negative.

9) Processing by Text-to-speech System

Speech researchers are increasingly focusing on the full range and variation of speech for signaling the social and psychological aspects of a message. It means that study of not just the conventional propositional content but also affective states [1]. The new generation of Text-To-Speech (TTS) systems should automatically deliver expressive cues when synthesizing an affective message [5], [6]. The recognization of affect in text is difficult problem. There are dificulties in translating human affect into explicit representations. One way is to presuppose the existence of some suitable taxonomy of affective states [6].Such taxonomy represents the nature of affective categories, which is generally described in two levels [9], [19], i.e., sentiment and emotion levels. Natural Language Processing (NLP) research community says that sentiment analysis classifies the problem into two classes that is positive and negative. Some authors also consider a neutral sentiment at the same level of the hierarchy [19]. TTS applications make use of this neutral sentiment to generate usual messages that are neither positive nor negative [15], [16].

B. Dimensionality Reduction and Weighting

In sentence-level Text Classification, the relative importance of features is of great relevance. But use of all the features together may increases the size of the feature space without providing satisfactory power [35]. Therefore, selecting and weighting the most relevant features raises the discriminating properties of the data, thus improving the classification effectiveness [27]. These methods are described as follows:

1) Term Selection:

It minimizes the dimensions of the feature space by removing the ones which do not contribute significantly to the task of classification . This is used to enhance both the effectiveness of the classifier and its computational performance given the fewer number of features to process.

2) Term Weighting:

It is used to raise the discriminating power of certain features without minimizing the dimensionality of the feature space.

Leung et al. suggest that semantical similarity may not imply sentimental similarity in sentiment analysis. Therefore they proposed a relative-frequency-based method to determine the sentiment orientation of an opinion. Their method estimates the sentiment orientation and opinion strength of a word with respect to a sentiment class as its relative frequency of appearance in that class. For example, if the word "excellent" appeared 9 times in *Positive* reviews and 1 times in *Negative* reviews, its strength with respect to *Positive* sentiment orientation is then 9/(9+1) = 0.9. (It means, no. of positive/(no. of positive + no. of negative)) Leung et al. deals with the rating inference problem, which classifies reviews with respect to rating scales. They hypothesized that some product features may be more important for determining the rating of a review, and therefore assign weights to opinions according to the estimated importance of their associated product features. They compute the weighted average sentiment orientation of opinions in a review, and then rate the review by mapping the weighted average onto an *n*-point rating scale.

C. Machine Learning

The machine learning approach is like that of the topic of classification, with the topics being sentiment classes such as Positive and Negative . It works by breaking down a review into words or phrases, representing the review as a document vector, and then classifying the reviews based on the document vectors.

` Pang et al. [27] investigated whether binary sentiment classification can be addressed using standard topic classification techniques. Three classifiers, including Naïve Bayes, Support Vector Machines (SVM) and Maximum Entropy are applied, to a movie review corpus by them. They also attempted to incorporate various features of the reviews into the standard bag-of- words model, such as the positions of words in the reviews, but the performance of the three classifiers was found inferior to those reported for topic classification. Pang and Lee concluded that sentiment classification is more difficult than topic classification, and that discourse analysis of reviews is necessary for more accurate sentiment analysis.

The rating inference as a metric-labeling problem is formulated by Pang and Lee. They first applied two *n*-ary classifiers, including one-vs-all (OVA) SVM and SVM regression, to classify reviews with respect to multi-point rating scales. They then use a metric-labeling algorithm to explicitly alter the results of the *n*-ary classifiers to ensure that similar items receive similar labels, determined using a similarity function. While term overlapping is a commonly-used similarity function in topic classification, it does not seem effective in identifying reviews having similar ratings [28]. They later proposed the Positive-Sentence Percentage (PSP) similarity function, computed as the number of positive sentences divided by the number of subjective sentences in a review. Experimental results in general show that using metric-labeling with PSP improves the performance of the *n*-ary classifiers.

Goldberg and Zhu later extended Pang and Lee's work using transductive semi-supervised learning. They demonstrated that unlabeled reviews can help improve classification accuracy [29].

Zhu and Goldberg proposed a kernel regression algorithm utilizing order preferences of unlabeled data, and successfully applied the algorithm to sentiment classification. The order preference of a pair of unlabeled data, x_i and x_j, indicates that x_i is preferred to x_j to some degree, even though the exact preferences for x_i and x_j are unknown. In the context of sentiment analysis, for example, given two reviews, one may be able to determine which review is more positive than the other without knowing the exact ratings associated with the reviews [30]. Zhu and Goldberg applied their algorithm to the rating inference problem, and showed that order preferences improved rating inference performance over standard regression.

4 Conclusion

Sentiment Analysis is a task of Natural language processing and information extraction which aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large number of documents. This paper discusses about taking the review of product using sentiment analysis and determining the polarity and features of the product based on the considered reviews

References

[1] N. Campbell, "Conversational speech synthesis and the need for some laughter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1171–1178, Jul. 2006.

[2] M. Skowron, H. Pirker, S. Rank, G. Paltoglou, J. Ahn, and S. Gobron, "No peanuts! Affective cues for the virtual bartender," in *Proc. FLAIRS '11*, May 2011.

[3] R. Calix, S. Mallepudi, B. Chen, and G. Knapp, "Emotion recognition in text for 3-D facial expression rendering," *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 544–551, Oct. 2010.

[4] T. Wilson and G. Hofer, "Using linguistic and vocal expressiveness in social role recognition," , pp. 419–422, 2011.

[5] F. Alías, X. Sevillano, J. C. Socoró, and X. Gonzalvo, "Towards high-quality next-generation text to-speech synthesis: A multidomain approach by automatic domain classification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 7, pp. 1340–1354, Sep. 2008.

[6] J. Bellegarda, "A data-driven affective analysis framework toward naturally expressive speech synthesis," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 5, pp. 1113–1122, Jul. 2011.

[7] J. Pitrelli, R. Bakis, E. Eide, R. Fernandez,W. Hamza, andM. Picheny, "The IBM expressive text-to-speech synthesis system for American English," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1099–1108, Jul. 2006.

[8] C. O. Alm, *Affect in Text and Speech*. Saarbrücken, Germany: VDM Verlag, 2009.

[9] V. Francisco, R. Hervás, F. Peinado, and P. Gervs, "EmoTales: Creating a corpus of folk tales with emotional annotations," *Lang. Res. Eval.*, vol. 45, pp. 1–41, Feb. 2011.

[10] M. Schröder, H. Pirker, M. Lamolle, F. Burkhardt, C. Peter, and E. Zovato, "Representing emotions and related states in technological systems," *Emotion-Oriented Systems – The Humaine Handbook*, pp. 367–386, 2011.

[11] G. O. Hofer, K. Richmond, and R. A. J. Clark, "Informed blending of databases for emotional speech synthesis," in *Proc. Interspeech '05*, Sep. 2005.

[12] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: Machine learning for text-based emotion prediction," in *Proc. HLT'05*, 2005, pp. 579–586, .

[13] A. R. F. Rebordao, M. A. M. Shaikh, K. Hirose, and N. Minematsu, "How to improve TTS systems for emotional expressivity," in *Proc. Interspeech'09*, Sep. 2009, pp. 524–527.

[14] M. A. M. Shaikh, A. R. F. Rebordao, K. Hirose, and M. Ishizuka,"Emotional speech synthesis by sensing affective information from text," in *Proc. ACII*, Sep. 2009, pp. 1–6.

[15] A. Trilla, F. Alías, and I. Lozano, "Text classification of domain-styled text and sentiment-styled text for expressive speech synthesis," in *Proc. VI Jornadas en Tecnología del Habla (FALA2010)*, Nov. 2010, pp. 75–78.

[16] A. Trilla and F. Alías, "Sentiment classification in english from sentence-level annotations of emotions regarding models of affect," in *Proc. Interspeech'09*, Sep. 2009, pp. 516–519.

[17] S. Tan and X. Cheng, "Improving SCL model for sentiment-transfer learning," in *Proc. NAACL-HLT*, 2009, pp. 181–184 .

[18] B. Pang and L. Lee, "opinionmining and sentiment analysis," *Foundat. Trends in Inf. Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[19] P. Shaver, J. Schwartz, D. Kirson, and C. O'Connor, "Emotion knowledge: Further exploration of a prototype approach," *J. Pers. Soc. Psychol.*, vol. 52, no. 6, pp. 1061–1086, 1987

[20] Lina Zhou, Pimwadee Chaovalit, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches", Proceedings of the 38th Hawaii International Conference on system sciences, 2005.

[21] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.

[22] Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack: 2003, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques", In: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003). Melbourne, Florida.

[23] Alekh Agarwal and Pushpak Bhattacharyya, "Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified", In Proceedings of the International Conference on Natural Language Processing (ICON), 2005.

[24] Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger, "Pulse: Mining Customer Opinions from Free Text Natural Language Processing", Microsoft Research, Redmond, WA 98052, USA.

[25] Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen. "Opinion Word Expansion and Target Extraction through Double Propagation." Computational Linguistics, March 2011, Vol. 37, No. 1: 9.27.

[26] Lei Zhang and Bing Liu. "Identifying Noun Product Features that Imply Opinions." *ACL-2011* (short paper), Portland, Oregon, USA, June 19-24, 2011.

[27] Sanjiv Das and Mike Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*, 2001.

[28] Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of WWW*, pages 519–528, 2003.

[29] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the Conference on Knowledge Capture (K-CAP)*, 2003.

[30] Erik Cambria, Robert Speer, Catherine Havasi, and Amir Hussain. Senticnet: A publicly available semantic resource for opinion mining. Artificial Intelligence, pages 14–18, 2010.

[31] Mostafa Karamibekr, Ali A. Ghorbani, Verb Oriented Sentiment Classification, 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

[32] Costin-Gabriel Chiru, Asmelash Teka Hadgu, Sentiment-Based Text Segmentation, 2013 2nd International Conference on Systems and Computer Science (ICSCS) Villeneuve d'Ascq, France, August 26-27, 2013

[33] Ms.K.Mouthami, Ms.K.Nirmala Devi, Dr.V.Murali Bhaskaran, Sentiment Analysis and Classification Based On Textual Reviews,

[34] Rabia Batool , Asad Masood Khattak , Jahanzeb Maqbool and Sungyoung Lee, Precise Tweet Classification and Sentiment Analysis

[35] Ellen Spertus. Smokey: Automatic recognition of hostile messages. In *Proceedings of Innovative Applications of Artificial Intelligence (IAAI)*, pages 1058–1065, 1997.

[36] Xin Jin, Ying Li, Teresa Mah, and Jie Tong. Sensitive webpage classification for content advertising. In *Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising*, 2007.

[37] Ahmed Abbasi. Affect intensity analysis of dark web forums. In *Proceedings of Intelligence and Security Informatics (ISI)*, pages 282–288, 2007.

[38] Miles Efron. Cultural orientation: Classifying subjective documents by cociation [sic] analysis. In 75 *Proceedings of the AAAI Fall Symposium on Style and Meaning in Language, Art, Music, and Design*, pages 41–48, 2004.

[39] Claire Cardie, Cynthia Farina, Thomas Bruce, and Erica Wagner. Using natural language processing to improve eRulemaking. In *Proceedings of Digital Government Research (dg.o)*, 2006.

[40] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 412–418, July 2004. Poster paper.

[41] Vasileios Hatzivassiloglou and Kathleen McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the Joint ACL/EACL Conference*, pages 174–181, 1997.

[42] F. Sebastiani, "Machine learning in automated text categorization,"

ACM Comput. Surv., vol. 34, pp. 1–47, 2002.
[43] U. D. Reichel and H. R. Pfitzinger, "Text preprocessing for speech synthesis," in *Proc. TC-STARWorkshop on Speech-to-Speech Translation*, Barcelona, Spain, Jun. 2006, pp. 207–212.
[44] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, MA: Cambridge Univ. Press, 2008.
[45] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980W.-K. Chen, *Linear Networks and Systems.* Belmont, Calif.: Wadsworth, pp. 123-135, 1993. (Book style)